Attorney's Docket No.: 07039-118002

APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: SULFOTRANSFERASE SEQUENCE VARIANTS

APPLICANT: RICHARD M. WEINSHILBOUM, REBECCA B. RAFTOGIANIS, THOMAS C. WOOD AND DIANE M. OTTERNESS

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No EL454377374US

I hereby certify under 37 CFR §1 10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D C 20231

April 10, 2001
Date of Deposit

Signature

Vince Defante
Typed or Printed Name of Person Signing Certificate

SULFOTRANSFERASE SEQUENCE VARIANTS

Technical Field

The invention relates to sulfotransferase nucleic acid sequence variants.

Statement as to Federally Sponsored Research

Funding for the work described herein was provided by the federal government, which has certain rights in the invention.

Background of the Invention

Pharmacogenetics is the study of the role of inheritance in variation of drug response, a variation that often results from individual differences in drug metabolism. Sulfation is an important pathway in the metabolism of many neurotransmitters, hormones, drugs and other xenobiotics. Sulfate conjugation is catalyzed by members of a gene superfamily of cytosolic sulfotransferase enzymes. It was recently agreed that "SULT" will be used as an abbreviation for these enzymes. These enzymes also are known as "PSTs" in the literature. Included among the nine cytosolic SULTs presently known to be expressed in human tissues are three phenol SULTs, SULT1A1, 1A2 and 1A3, which catalyze the sulfate conjugation of many phenolic drugs and other xenobiotics.

Biochemical studies of human phenol SULTs led to the identification of two isoforms that were defined on the basis of substrate specificities, inhibitor sensitivities and thermal stabilities. A thermostable (TS), or phenol-preferring form, and a thermolabile (TL), or monoamine-preferring form, were identified. "TS PST" preferentially catalyzed the sulfation at micromolar concentrations of small planar phenols such as 4-nitrophenol and was sensitive to inhibition by 2,6-dichloro-4-nitrophenol (DCNP). "TL PST" preferentially catalyzed the sulfation of micromolar concentration of phenolic monoamines such as dopamine and was relatively insensitive

to DCNP inhibition. Weinshilboum, R.M. Fed. Proc., 45:2223 (1986). Both of these biochemically-defined activities were expressed in a variety of human tissues including liver, brain, jejunum and blood platelets. Human platelet TS PST displayed wide individual variations, not only in level of activity, but also in thermal stability. Segregation analysis of data from family studies of human platelet TS PST showed that levels of this activity as well as individual variations in its thermal stability were controlled by genetic variation. Price, P.A. et al., Genetics, 122:905-914 (1989).

Molecular genetic experiments indicated that there are three "PST genes" in the human genome, two of which, *SULT1A1* (*STP1*) and *SULT1A2* (*STP2*), encode proteins with TS PST-like activity, *SULT1A1* (*TS PST1*) and *SULT1A2* (*TS PST2*), respectively. The remaining gene, *SULT1A3* (*STM*), encodes a protein with TL PST-like activity, *SULT1A3* (*TL PST*). DNA sequences and structures of the genes for these enzymes are highly homologous, and all three map to a phenol SULT gene complex on the short arm of human chromosome 16. Weinshilboum, R. et al., FASEB J., 11(1):3-14 (1997).

Summary of the Invention

The invention is based on the discovery of several common *SULT1A1* and *SULT1A2* alleles encoding enzymes that differ functionally and are associated with individual differences in phenol SULT properties in platelets and liver. In addition, the invention is based on the discovery of *SULT1A3* sequence variants. These discoveries permit use of SULT genomic and biochemical pharmacogenetic data to better understand the possible contribution of inheritance to individual differences in the sulfate conjugation of drugs and other xenobiotics in humans. Thus, the identification of SULT allozymes and alleles allows sulfonator status of a subject to be assessed. The information and insight obtained thereby allows tailoring of particular treatment regimens in the subject. In addition, risk estimates for hormone dependent diseases can be determined.

The invention features an isolated nucleic acid molecule including a *SULT1A3* nucleic acid sequence. The sulfotransferase nucleic acid sequence includes a nucleotide sequence variant and nucleotides flanking the sequence variant. A nucleic acid construct that includes such sulfotransferase nucleic acid sequences is also described. The *SULT1A3* sulfotransferase nucleic acid sequence can encode a sulfotransferase polypeptide including an amino acid sequence variant. *SULT1A3* nucleotide sequence variants can be within an intron. For example, introns 4 and 6 each can include an adenine at nucleotide 69. Intron 7 can include a thymine at nucleotide 113. *SULT1A3* nucleotide sequence variants can include insertion of nucleotides within intron sequences. The nucleotide sequence 5'-CAGT-3' can be inserted, for example, within intron 3. A *SULT1A3* nucleotide sequence variant also can include a guanine at nucleotide 105 of the coding sequence.

The invention also features *SULT1A1* and *SULT1A2* nucleotide sequence variants. The *SULT1A1* nucleotide sequence variants can include, for example, a cytosine at nucleotide 138 of intron 1A or a thymine at nucleotide 34 of intron 5. A *SULT1A1* variant also can include, for example, an adenine at nucleotide 57, 110, or 645 of the *SULT1A1* coding sequence. The *SULT1A1* nucleic acid sequence can encode a sulfotransferase polypeptide having, for example, a glutamine at amino acid 37. *SULT1A2* nucleotide sequence variants can include a thymine at nucleotide 78 of intron 5 or a thymine at nucleotide 9 of intron 7. The coding sequence of *SULT1A2* can include a thymine of nucleotide 550. The *SULT1A2* nucleic acid sequence can encode, for example a cysteine at amino acid 184.

In another aspect, the invention features a method for determining a risk estimate of a hormone disease in a patient. The method includes detecting the presence or absence of a sulfotransferase nucleotide sequence variant in a patient, and determining the risk estimate based, at least in part, on presence or absence of the variant in the patient. The hormone dependent disease can be, for example, breast cancer, prostate cancer or ovarian cancer.

The invention also features a method for determining sulfonator status in a subject. The method includes detecting the presence or absence of a sulfotransferase allozyme or nucleotide sequence variant in a subject, and determining the sulfonator status based, at least in part, on said determination.

An antibody having specific binding affinity for a sulfotransferase polypeptide is also described.

The invention also features isolated nucleic acid molecules that include a sulfotransferase nucleic acid sequence that encode a sulfotransferase allozyme. The allozyme can be selected from the group consisting of *SULT1A1*4*, *SULT1A2*4*, *SULT1A2*5*, and *SULT1A2*6*. Sulfotransferase nucleic acid sequences that include sulfotransferase alleles selected from the group consisting of *SULT1A1*1*, *SULT1A1*2*, *SULT1A1*3A*, *SULT1A1*3B* and *SULT1A1*4* also are featured. In particular, the *SULT1A1*1* allele can be *SULT1A1*1A* to *SULT1A1*1K*. The *SULT1A2* allele can be *SULT1A2*1A-1D*, *SULT1A2*2A-2C*, *SULT1A2*3A-3C* or *SULT1A2*4-*6*.

The invention also relates to an article of manufacture that includes a substrate and an array of different sulfotransferase nucleic acid molecules immobilized on the substrate. Each of the different sulfotransferase nucleic acid molecules includes a different sulfotransferase nucleotide sequence variant and nucleotides flanking the sequence variant. The array of different sulfotransferase nucleic acid molecules can include at least two nucleotide sequence variants of *SULT1A1*, *SULT1A2*, or *SULT1A3*.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In

addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

Figure 1 represents human platelet TS PST phenotypes. Figure 1A is a scattergram that depicts the relationship between TS PST enzymatic activity and thermal stability in 905 human platelet samples. Figure 1B is a scattergram that correlates human platelet *SULT1A1* genotype with TS PST phenotype.

Figure 2 is a representation of human *SULT1A1*, *SULT1A2*, and *SULT1A3* gene structures and the PCR strategy used to amplify the open reading frame (ORF) of each gene in three segments. Black rectangles represent exons that encode cDNA ORF sequence, while open rectangles represent exon or portions of exons that encode cDNA untranslated region (UTR) sequence. Roman numerals are exon numbers, and arabic numerals are exon lengths in bp. Gene lengths in kb from initial to final exons are also indicated. Forward and reverse arrows indicate the placement within introns of the PCR primers used to amplify, in three separate reactions, the ORFs of *SULT1A1* and *SULT1A2*.

Figure 3 is a scattergram that depicts the relationship between TS PST enzymatic activity and thermal stability in 61 human liver biopsy samples.

Figures 4A and 4B are scattergrams that depict the correlation of *SULT1A1* and *SULT1A2* genotypes with human liver TS PST phenotype. TS PST phenotypes in the human liver samples depicted as in Fig. 3 are shown with (A) common *SULT1A1* allozymes or (B) common *SULT1A2* allozymes superimposed. In (B) three samples are not shown because they contain *SULT1A2* allozymes that were observed only once in this population sample.

Figure 5 is the gene sequence of *SULT1A1* (SEQ ID NO:29).

Figure 6 is the gene sequence of *SULT1A2* (SEQ ID NO:31).

Figure 7 is the gene sequence of *SULT1A3* (SEQ ID NO:33).

Detailed Description

The invention features an isolated nucleic acid molecule that includes a sulfotransferase nucleic acid sequence. The sulfotransferase nucleic acid sequence includes a nucleotide sequence variant and nucleotides flanking the sequence variant. As used herein, "isolated nucleic acid" refers to a sequence corresponding to part or all of the sulfotransferase gene, but free of sequences that normally flank one or both sides of the sulfotransferase gene in a mammalian genome. The term "sulfotransferase nucleic acid sequence" refers to a nucleotide sequence of at least about 14 nucleotides in length. For example, the sequence can be about 14 to 20, 20-50, 50-100 or greater than 100 nucleotides in length. Sulfotransferase nucleic acid sequences can be in sense or antisense orientation. Suitable sulfotransferase nucleic acid sequences include *SULT1A1*, *SULT1A2* and *SULT1A3* nucleic acid sequences. As used herein, "nucleotide sequence variant" refers to any alteration in the wild-type gene sequence, and includes variations that occur in coding and non-coding regions, including exons, introns, promoters and untranslated regions.

In some instances, the nucleotide sequence variant results in a sulfotransferase polypeptide having an altered amino acid sequence. The term "polypeptide" refers to a chain of at least four amino acid residues. Corresponding sulfotransferase polypeptides, irrespective of length, that differ in amino acid sequence are herein referred to as allozymes. For example, a sulfotransferase nucleic acid sequence can be a *SULT1A1* nucleic acid sequence and include an adenine at nucleotide 110. This nucleotide sequence variant encodes a sulfotransferase polypeptide having a glutamine at amino acid residue 37. This polypeptide would be considered an allozyme with respect to a corresponding sulfotransferase polypeptide having an arginine at amino acid residue 37. In addition, the nucleotide variant can include an adenine at nucleotide 638 or a guanine at nucleotide 667, and encode a

sulfotransferase polypeptide having a histidine at amino acid residue 213 or a valine at amino acid residue 223, respectively.

As described herein, there are at least four *SULT1A1* allozymes. *SULT1A1*1* is the most common and contains an arginine at residues 37 and 213, and a methionine at residue 223. *SULT1A1*2* contains an arginine at residue 37, a histidine at residue 213 and a methionine at residue 223. *SULT1A1*3* contains an arginine at residues 37 and 213, and a valine at residue 223. *SULT1A*4* is the least common, and contains a glutamine at residue 37, an arginine at residue 213, and a methionine at residue 223.

The sulfotransferase nucleic acid sequence also can encode SULT1A2 polypeptide variants. Non-limiting examples of SULT1A2 polypeptide variants include an isoleucine at amino acid residue 7, a leucine at amino acid residue 19, a cysteine at amino acid residue 184, or a threonine at amino acid 235. These polypeptide variants are encoded by nucleotide sequence variants having a cytosine at nucleotide 20, a thymine at nucleotide 56, a thymine at nucleotide 50 and a cytosine at nucleotide 704.

There are at least six different SULT1A2 allozymes that differ at residues 7, 19, 184 and 235. For example, SULT1A2*1 contains an isoleucine, a proline, an arginine and an asparagine at residues 7, 19, 184 and 235, respectively, and represents the most common allozyme. SULT1A2*2 differs from SULT1A2*1 in that it contains a threonine at residues 7 and 235. SULT1A2*3 differs from SULT1A2*1 in that it contains a leucine at residue 19. SULT1A2*4 differs from SULT1A2*2 in that it contains a cysteine at residue 184. SULT1A2*5 differs from SULT1A2*1 in that it contains a threonine at residue 7. SULT1A2*6 differs from SULT1A2*1 in that it contains an isoleucine at residue 7.

As described herein, SULT1A1*2 and SULT1A2*2 are associated with decreased TS PST thermal stability in the human liver, but the biochemical and physical properties of recombinant SULT allozymes indicated that the "TS PST

phenotype" in the liver is most likely due to expression of SULT1A1. For example, based both on its apparent K_m value for 4-nitrophenol and its T_{50} value, SULT1A1*2 was not consistently associated with low levels of TS PST activity in the liver, but was uniformly associated with decreased levels of platelet TS PST activity and thermal stability. It appears that SULT1A1*2 is associated with lower levels of TS PST activity in tissue from subjects with benign rather than neoplastic disease.

Certain sulfotransferase nucleotide variants do not alter the amino acid sequence. Such variants, however, could alter regulation of transcription as well as mRNA stability. *SULT1A1* variants can occur in intron sequences, for example, within intron 1A and introns 5-7 (i.e., intron 5 is immediately after exon 5 in Figure 5). In particular, the nucleotide sequence variant can include a cytosine at nucleotide 138 of intron 1A, or a thymine at nucleotide 34 or an adenine at nucleotide 35 of intron 5. Intron 6 sequence variants can include a guanine at nucleotide 11, a cytosine at nucleotide 17, an adenine at nucleotide 35, a guanine at nucleotide 45, a guanine at nucleotide 64, a cytosine at nucleotide 488, and an adenine at nucleotide 509. Intron 7 variants can include a thymine at nucleotide 17, a cytosine at nucleotide 69 and a guanine at nucleotide 120. *SULT1A1* nucleotide sequence variants that do not change the amino acid sequence also can be within an exon or in the 3' untranslated region. For example, the coding sequence can contain an adenine at nucleotide 57, a cytosine at nucleotide 153, a guanine at nucleotide 162, a cytosine at nucleotide 600, or an adenine at nucleotide 645. The 3' untranslated region can contain a guanine at nucleotide 902 or a thymine at nucleotide 973.

Similarly, certain *SULT1A2* and *SULT1A3* variants do not alter the amino acid sequence. Such *SULT1A2* nucleotide sequence variants can be within an intron sequence, a coding sequence or within the 3' untranslated region. In particular, the nucleotide variant can be within intron 2, 5 or 7. For example, intron 2 can contain a cytosine at nucleotide 34. Intron 5 can include a thymine at nucleotide 78, and intron 7 can include a thymine at nucleotide 9. In addition, a cytosine can be at nucleotide

24 or a thymine at nucleotide 895 in *SULT1A2* coding sequence. A guanine can be at nucleotide 902 in the 3' untranslated region. *SULT1A3* nucleotide sequences variant can include a guanine at nucleotide 105 of the coding region (within exon 3). In addition, intron 3 of *SULT1A3* can include an insertion of nucleotides. For example, the four nucleotides 5'-CAGT-3' can be inserted between nucleotides 83 and 84 of intron 3. Introns 4, 6, and 7 also can contain sequence variants. For example, nucleotide 69 of introns 4 and 6 can contain an adenine. Nucleotide 113 of intron 7 can contain a thymine.

Sulfotransferase allozymes as described above are encoded by a series of sulfotransferase alleles. These alleles represent nucleic acid sequences containing sequence variants, typically multiple sequence variants, within intron, exon and 3' untranslated sequences. Representative examples of single nucleotide variants are described above. Table 3 sets out a series of 13 *SULT1A1* alleles (*SULT1A1*1A* to *SULT1A1*1K*) that encode *SULT1A1*1*. *SULT1A1*1A* to *SULT1A1*1K* range in frequency from about 0.7% to about 33%, as estimated from random blood donors and hepatic biopsy samples. Two alleles, *SULT1A1*3A* and *SULT1A1*3B* each encode *SULT1A1*3*, and represent about 0.3% to about 1.6% of all *SULT1A1* alleles. *SULT1A1*2* and *SULT1A1*4* are encoded by single alleles, *SULT1A1*2* and *SULT1A1*4*, respectively. *SULT1A1*2* represents about 31% of the alleles, whereas *SULT1A1*4* accounts for only about 0.3% of the alleles.

Numerous *SULT1A2* alleles also exist (Table 2A). For example, *SULT1A2*1* is encoded by four alleles (*SULT1A2*1A* to *SULT1A2*1D*) that range in frequency from 0.8% to about 47%. *SULT1A2*2* and *SULT1A2*3* are each encoded by three alleles (**2A* - **2C* and **3A* - **3C*). These alleles range in frequency from 0.8% up to about 26%. Single alleles encode *SULT1A2*4*, *SULT1A2*5*, and *SULT1A2*6*, with each representing about 0.8% of the *SULT1A2* alleles. As described herein, *SULT1A2* alleles are in linkage disequilibrium with the alleles for *SULT1A1*.

The relatively large number of alleles and allozymes for *SULT1A1* and *SULT1A2*, with three common allozymes for each gene, indicates the potential complexity of SULT pharmacogenetics. Such complexity emphasizes the need for determining single nucleotide variants, as well as complete haplotypes of patients. For example, an article of manufacture that includes a substrate and an array of different sulfotransferase nucleic acid molecules immobilized on the substrate allows complete haplotypes of patients to be assessed. Each of the different sulfotransferase nucleic acid molecules includes a different sulfotransferase nucleotide sequence variant and nucleotides flanking the sequence variant. The array of different sulfotransferase nucleic acid molecules can include at least two nucleotide sequence variants of *SULT1A1*, *SULT1A2*, or *SULT1A3*, or can include all of the nucleotide sequence variants known for each gene.

Suitable substrates for the article of manufacture provide a base for the immobilization of nucleic acid molecules into discrete units. For example, the substrate can be a chip or a membrane. The term "unit" refers to a plurality of nucleic acid molecules containing the same nucleotide sequence variant. Immobilized nucleic acid molecules are typically about 20 nucleotides in length, but can vary from about 14 nucleotides to about 100 nucleotides in length. In practice, a sample of DNA or RNA from a subject can be amplified, hybridized to the article of manufacture, and then hybridization detected. Typically, the amplified product is labeled to facilitate hybridization detection. See, for example, Hacia, J.G. et al., Nature Genetics, 14:441-447 (1996); and U.S. Patent Nos. 5,770,722 and 5,733,729.

As a result of the present invention, it is now possible to determine sulfonator status of a subject. As used herein "sulfonator status" refers to the ability of a subject to transfer a sulfate group to a substrate. A variety of drugs (e.g., acetaminophen), hormones (e.g., estrogen) and neurotransmitters (e.g., dopamine and other phenolic monoamines) are substrates for these enzymes. Generally, sulfonation is considered a detoxification mechanism, as reaction products are more readily

excreted. Certain substrates, however, become more reactive upon sulfonation. For example, the N-hydroxy metabolite of 2-acetylaminoflourene is converted to a N-O-sulfate ester, which is reactive with biological macromolecules. Thus, a determination of the presence or absence of nucleotide sequence variants or allozymes facilitates the prediction of therapeutic efficacy and toxicity of drugs on an individual basis, as well as the ability to biotransform certain hormones and neurotransmitters. In addition, the ability to sulfonate hormones may play a role in cancer.

The presence or absence of sulfotransferase variants allows the determination of a risk estimate for the development of a hormone dependent disease. As used herein, "hormone dependent disease" refers to a disease in which a hormone plays a role in the pathophysiology of the disease. Non-limiting examples of hormone dependent diseases include breast cancer, ovarian cancer, and prostate cancer. Risk estimate indicates the relative risk a subject has for developing a hormone dependent disease. For example, a risk estimate for development of breast cancer can be determined based on the presence or absence of sulfotransferase variants. A subject containing, for example, the *SULT1A1*2*, of sulfotransferase variant may have a greater likelihood of having breast cancer. Additional risk factors include, for example, family history of breast cancer and other genetic factors such as mutations within the BRCA1 and BRCA2 genes.

Sulfotransferase nucleotide sequence variants can be assessed, for example, by sequencing exons and introns of the sulfotransferase genes, by performing allele-specific hybridization, allele-specific restriction digests, mutation specific polymerase chain reactions (MSPCR), or by single-stranded conformational polymorphism (SSCP) detection. Polymerase chain reaction (PCR) refers to a procedure or technique in which target nucleic acids are amplified. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. PCR can be used to amplify specific

sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA. Primers are typically 14 to 40 nucleotides in length, but can range from 10 nucleotides to hundreds of nucleotides in length. PCR is described, for example in PCR Primer: A Laboratory Manual, Ed. by Dieffenbach, C. and Dveksler, G., Cold Spring Harbor Laboratory Press, 1995. Nucleic acids also can be amplified by ligase chain reaction, strand displacement amplification, self-sustained sequence replication or nucleic acid sequence-based amplification. See, for example, Lewis, R. Genetic Engineering News, 12(9):1 (1992); Guatelli et al., Proc. Natl. Acad. Sci. USA, 87:1874-1878 (1990); and Weiss, R., Science, 254:1292 (1991).

Genomic DNA is generally used in the analysis of sulfotransferase nucleotide sequence variants. Genomic DNA is typically extracted from peripheral blood samples, but can be extracted from such tissues as mucosal scrapings of the lining of the mouth or from renal or hepatic tissue. Routine methods can be used to extract genomic DNA from a blood or tissue sample, including, for example, phenol extraction. Alternatively, genomic DNA can be extracted with kits such as the QIAamp® Tissue Kit (Qiagen, Chatsworth, CA), Wizard® Genomic DNA purification kit (Promega, Madison, WI) and the A.S.A.P.™ Genomic DNA isolation kit (Boehringer Mannheim, Indianapolis, IN).

For example, exons and introns of the sulfotransferase gene can be amplified through PCR and then directly sequenced. This method can be varied, including using dye primer sequencing to increase the accuracy of detecting heterozygous samples. Alternatively, a nucleic acid molecule can be selectively hybridized to the PCR product to detect a gene variant. Hybridization conditions are selected such that the nucleic acid molecule can specifically bind the sequence of interest, e.g., the variant nucleic acid sequence. Such hybridizations typically are performed under high stringency as some sequence variants include only a single nucleotide difference. High stringency conditions can include the use of low ionic strength solutions and high temperatures for washing. For example, nucleic acid

molecules can be hybridized at 42°C in 2X SSC (0.3M NaCl/0.03 M sodium citrate/0.1% sodium dodecyl sulfate (SDS) and washed in 0.1X SSC (0.015M NaCl/0.0015 M sodium citrate), 0.1% SDS at 65°C. Hybridization conditions can be adjusted to account for unique features of the nucleic acid molecule, including length and sequence composition.

Allele-specific restriction digests can be performed in the following manner. For example, if a nucleotide sequence variant introduces a restriction site, restriction digest with the particular restriction enzyme can differentiate the alleles. For *SULT1* variants that do not alter a common restriction site, primers can be designed that introduce a restriction site when the variant allele is present, or when the wild-type allele is present. For example, the *SULT1A*2* allele does not have an altered restriction site. A *KasI* site can be introduced in all *SULT1A1* alleles, except *SULT1A*2*, using a mutagenic primer (e.g., 5'CCA CGG TCT CCT CTG GCA GGG GG 3', SEQ ID NO:1). A portion of *SULT1A1* alleles can be amplified using the mutagenic primer and a primer having, for example, the nucleotide sequence of 5' GTT GAG GAG TTG GCT CTG CAG GGT C 3' (SEQ ID NO:2). A *KasI* digest of *SULT1A1* alleles, other than *SULT1A*2*, yield restriction products of about 173 base pairs (bp) and about 25 bp. In contrast, the *SULT1A*2* allele is not cleaved, and thus yields a restriction product of about 198 bp.

The *SULT1A2*2* allele can be detected using a similar strategy. For example, an additional *StyI* site can be introduced in the *SULT1A2*2* allele using the mutagenic primer 5' CAC GTA CTC CAG TGG CGG GCC CTA G 3' (SEQ ID NO:3). Upon amplification of a portion of the *SULT1A2* alleles using the mutagenic primer and a primer having the nucleotide sequence of 5' GGA ACC ACC ACA TTA GAA C 3' (SEQ ID NO:4), a *StyI* digest yields restriction products of 89 bp, 119 bp and 25 bp for *SULT1A2*2*. The other *SULT1A2* alleles described herein yield restriction products of 89 bp and 144 bp.

Certain variants, such as the insertion within intron 3 of the *SULT1A3* gene discussed above, change the size of the DNA fragment encompassing the variant. The insertion of nucleotides can be assessed by amplifying the region encompassing the variant and determining the size of the amplified products in comparison with size standards. For example, the intron 3 region of the *SULT1A3* gene can be amplified using a primer set from either side of the variant. One of the primers is typically labeled, for example, with a fluorescent moiety, to facilitate sizing. The amplified products can be electrophoresed through acrylamide gels using a set of size standards that are labeled with a fluorescent moiety that differs from the primer.

PCR conditions and primers can be developed that amplify a product only when the variant allele is present or only when the wild-type allele is present (MSPCR or allele-specific PCR). For example, patient DNA and a control can be amplified separately using either a wild-type primer or a primer specific for the variant allele. Each set of reactions is then examined for the presence of amplification products using standard methods to visualize the DNA. For example, the reactions can be electrophoresed through an agarose gel and DNA visualized by staining with ethidium bromide or other DNA intercalating dye. In DNA samples from heterozygous patients, reaction products would be detected in each reaction. Patient samples containing solely the wild-type allele would have amplification products only in the reaction using the wild-type primer. Similarly, patient samples containing solely the variant allele would have amplification products only in the reaction using the variant primer.

Mismatch cleavage methods also can be used to detect differing sequences by PCR amplification, followed by hybridization with the wild-type sequence and cleavage at points of mismatch. Chemical reagents, such as carbodiimide or hydroxylamine and osmium tetroxide can be used to modify mismatched nucleotides to facilitate cleavage.

Alternatively, sulfotransferase nucleotide sequence variants can be detected by antibodies that have specific binding affinity for variant sulfotransferase polypeptides. Variant sulfotransferase polypeptides can be produced in various ways, including recombinantly. The genomic nucleic acid sequences of *SULT1A1*, *SULT1A2* and *SULT1A3* have GenBank accession numbers of U52852, U34804 and U20499, respectively. Amino acid changes can be introduced by standard techniques including oligonucleotide-directed mutagenesis and site-directed mutagenesis through PCR. See, Short Protocols in Molecular Biology, Chapter 8, Green Publishing Associates and John Wiley & Sons, Edited by Ausubel, F.M et al., 1992.

A nucleic acid sequence encoding a sulfotransferase variant polypeptide can be ligated into an expression vector and used to transform a bacterial or eukaryotic host cell. In general, nucleic acid constructs include a regulatory sequence operably linked to a sulfotransferase nucleic acid sequence. Regulatory sequences do not typically encode a gene product, but instead affect the expression of the nucleic acid sequence. In bacterial systems, a strain of *Escherichia coli* such as BL-21 can be used. Suitable *E. coli* vectors include the pGEX series of vectors that produce fusion proteins with glutathione S-transferase (GST). Transformed *E. coli* are typically grown exponentially, then stimulated with isopropylthiogalactopyranoside (IPTG) prior to harvesting. In general, such fusion proteins are soluble and can be purified easily from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In eukaryotic host cells, a number of viral-based expression systems can be utilized to express sulfotransferase variant polypeptides. A nucleic acid encoding a sulfotransferase variant polypeptide can be cloned into, for example, a baculoviral vector and then used to transfect insect cells. Alternatively, the nucleic acid encoding

a sulfotransferase variant can be introduced into a SV40, retroviral or vaccinia based viral vector and used to infect host cells.

Mammalian cell lines that stably express sulfotransferase variant polypeptides can be produced by using expression vectors with the appropriate control elements and a selectable marker. For example, the eukaryotic expression vector pCR3.1 (Invitrogen, San Diego, CA) is suitable for expression of sulfotransferase variant polypeptides in, for example, COS cells. Following introduction of the expression vector by electroporation, DEAE dextran, or other suitable method, stable cell lines are selected. Alternatively, amplified sequences can be ligated into a mammalian expression vector such as pcDNA3 (Invitrogen, San Diego, CA) and then transcribed and translated *in vitro* using wheat germ extract or rabbit reticulocyte lysate.

Sulfotransferase variant polypeptides can be purified by known chromatographic methods including DEAE ion exchange, gel filtration and hydroxylapatite chromatography. Van Loon, J.A. and R.M. Weinshilboum, Drug Metab. Dispos., 18:632-638 (1990); Van Loon, J.A. et al., Biochem. Pharmacol., 44:775-785 (1992).

Various host animals can be immunized by injection of a sulfotransferase variant polypeptide. Host animals include rabbits, chickens, mice, guinea pigs and rats. Various adjuvants that can be used to increase the immunological response depend on the host species and include Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin and dinitrophenol. Polyclonal antibodies are heterogenous populations of antibody molecules that are contained in the sera of the immunized animals. Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be prepared using a sulfotransferase variant polypeptide and standard hybridoma technology. In particular, monoclonal antibodies can be obtained

by any technique that provides for the production of antibody molecules by continuous cell lines in culture such as described by Kohler, G. et al., Nature, 256:495 (1975), the human B-cell hybridoma technique (Kosbor et al., Immunology Today, 4:72 (1983); Cole et al., Proc. Natl. Acad. Sci USA, 80:2026 (1983)), and the EBV-hybridoma technique (Cole et al., "Monoclonal Antibodies and Cancer Therapy", Alan R. Liss, Inc., pp. 77-96 (1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the monoclonal antibodies of the invention can be cultivated *in vitro* and *in vivo*.

Antibody fragments that have specific binding affinity for a sulfotransferase variant polypeptide can be generated by known techniques. For example, such fragments include but are not limited to $F(ab')_2$ fragments that can be produced by pepsin digestion of the antibody molecule, and Fab fragments that can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed. See, for example, Huse et al., Science, 246:1275 (1989). Once produced, antibodies or fragments thereof are tested for recognition of sulfotransferase variant polypeptides by standard immunoassay methods including ELISA techniques, radioimmunoassays and Western blotting. See, Short Protocols in Molecular Biology, Chapter 11, Green Publishing Associates and John Wiley & Sons, Edited by Ausubel, F.M et al., 1992.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

Examples

1.0 Methods and Materials

1.1 Tissue Samples

Human hepatic "surgical waste" tissue was obtained from 61 patients undergoing clinically-indicated hepatectomies or open hepatic biopsies and was stored at -80°C. These frozen hepatic tissue samples were homogenized in 5 mM potassium

phosphate buffer, pH 6.5, and centrifuged at 100,000 x g for 1 hr to obtain high-speed supernatant (HSS) cytosolic preparations. Campbell, N.R.C. et al., Biochem. Pharmacol., 36:1435-1446 (1987). Platelet samples were obtained from blood samples from 905 members of 134 randomly selected families at the Mayo Clinic in Rochester, MN. All tissue samples were obtained under guidelines approved by the Mayo Clinic Institutional Review Board.

1.2 PST Enzyme Activity, Thermal Stability and Inhibitor Sensitivity

TS PST enzyme activity was measured with an assay that involves the sulfate conjugation of substrate, in this case 4-nitrophenol, in the presence of [^{35}S]-3′-phosphoadenosine-5′-phosphosulfate (PAPS), the sulfate donor for the reaction. See, Campbell, N.R.C. et al., Biochem. Pharmacol., 36:1435-1446 (1987). Blanks were samples that did not contain sulfate acceptor substrate. Unless otherwise stated, concentrations of 4-nitrophenol and PAPS were 4 μM and 0.4 μM, respectively. Substrate kinetic experiments were conducted in the presence of a series of concentrations of 4-nitrophenol and PAPS to make it possible to calculate apparent K_m values. Enzyme activity was expressed as nmoles of sulfate conjugated product formed per hr of incubation. Protein concentrations were measured by the dye-binding method of Bradford with bovine serum albumin (BSA) as a standard.

Enzyme thermal stability was determined as described by Reiter and Weinshilboum, Clin. Pharmacol. Ther., 32:612-621 (1982). Specifically, hepatic HSS preparations or platelet preparations were thawed, diluted and were then either subjected to thermal inactivation for 15 min at 44°C or were kept on ice as a control. In these experiments, heated over control (H/C) ratios were used as a measure of thermal stability. The thermal stability of recombinant proteins was measured by incubating diluted, transfected COS-1 cell HSS for 15 min in a Perkin Elmer 2400 thermal cycler at a series of temperatures. All samples were placed on ice immediately after the thermal inactivation step, and PST activity was measured in both heated and control samples. Thermal inactivation curves were then constructed

for each recombinant protein by plotting SULT activity expressed as a percentage of the control value. The concentration of 4-nitrophenol used to assay each of the recombinant proteins was determined on the basis of the results of the substrate kinetic experiments during which apparent K_m values had been determined. Those concentrations were: *SULT1A1* (*1, *2, *3), 4 μM; *SULT1A2*, 100 μM; *SULT1A2*2*, 3 mM; *SULT1A2*3*, 50 μM; and *SULT1A3*, 3 mM.

DCNP inhibition was determined by measuring enzyme activity in the presence of a series of DCNP concentrations dissolved in dimethylsulfoxide. Blank samples for those experiments contained the appropriate concentration of DCNP, but no sulfate acceptor substrate. The concentration of 4-nitrophenol used to study each recombinant protein was the same as was used in the thermal stability experiments. All assays for the determination of apparent K_m values, thermal stability or DCNP inhibition were performed in triplicate, and all experiments were performed at least three times, i.e., each of the data points shown subsequently represents the average of at least nine separate assays.

1.3 PCR Amplification and DNA Sequencing

Total genomic DNA was isolated from the human liver biopsy samples with a QIAamp Tissue Kit (Qiagen, Inc., Chatsworth, CA). In addition, genomic DNA was isolated from 150 randomly selected Caucasian blood donors at the Mayo Clinic Blood Blank. Gene-specific primers for the PCR were designed by comparing the sequences of *SULT1A1*, *SULT1A2*, and *SULT1A3* (Genbank accession numbers U52852, U34804 and U20499, respectively) and identifying intron sequences that differed among the three genes. These gene-specific primers were then used to amplify, in three separate segments for each gene, the coding regions of either *SULT1A1* or *SULT1A2* (Fig. 2). To assure specificity, an initial long PCR amplification was performed using oligonucleotide primers that annealed to unique sequences present in the 5′-and 3′-flanking regions of each gene. Those long PCR products were then used as templates for the subsequent PCR reactions to amplify

coding regions of the genes. Sequences of the PCR primers used to perform these experiments are listed in Table 1. In Table 1, "I" represents "intron", "F" represents "forward", "R" represents "reverse" and "D" ("downstream") represents 3′-flanking region of the gene.

DNA sequencing was performed with single-stranded DNA as template to help assure the detection of heterozygous samples. To make that possible, single-stranded DNA was generated by exonuclease digestion of either the sense or antisense strand of the double-stranded PCR amplification products. Phosphorothioate groups were conjugated to the 5′-end of either the forward or reverse PCR primer, depending on which of the two strands was to be protected from exonuclease digestion. Specifically, the PCR amplification of gene segments was performed in a 50 μl reaction mixture using Amplitaq Gold DNA polymerase (Perkin Elmer). Digestion of the non-phosphorothioated strand involved incubation of 16 μl of the post-amplification reaction mixture with 20 units of T7 gene 6 exonuclease (United States Biochemical, Cleveland, OH) in 10 mM Tris-HCl buffer, pH 7.5, containing 200 μM DTT and 20 μg/ml BSA. This mixture was incubated at 37°C for 4 hr, followed by inactivation of the exonuclease by incubation at 80°C for 15 min. The resulting single stranded DNA was used as a sequencing template after PCR primers and salts had been removed with a Microcon-100 microconcentrator (Amicon, Beverly, MA). DNA sequencing was performed in the Mayo Clinic Molecular Biology Core Facility with an ABI Model 377 sequencer (Perkin Elmer, Foster City, CA) using dye terminator cycler sequencing chemistry.

TABLE 1
PCR Primers

REACTION	PRIMER	Seq ID	PRIMER SEQUENCE (5' to 3')
***SULT1A1* Gene-Specific Amplifications**			
Long PCR	1AF(-119) DR3296	5 6	CCTGGAGACCTTCACACACCCTGATA CCACTCTGCCTGGCCCACAATCATA
Segment 1	I1AF11 I4R83	7 8	GCTGGGGAACCACCGCATTAGAG AACTCCCAACCTCACGTGATCTG
Segment 2	I4F1018 I6R93	9 10	CCTCAGGTTCCTCCTTTGCCAAT TGCCAAGGGAGGGGGCTGGGTGA
Segment 3	I6F395 DR3296	11 12	GTTGAGGAGTTGGCTCTGCAGGGTC CCACTCTGCCTGGCCCACAATCATA
***SULT1A2* Gene-Specific Amplifications**			
Long PCR	1AF(-90) DR4590	13 14	GGGCCCCGTTCCACGAGGGTGCTTTCAC TGACCCCACTAGGAAGGGAGTCAGCACCCCTACT
Segment 1	I1AF16 I4R86	15 16	GGAACCACCACATTAGAAC TGGAACTTCTGGCTTCAAGGGATCT
Segment 2	I4F1117 I6R81	17 18	CCTCAGCTTCCTCCTTTGCCAAA TGGCTGGGTGGCCTTGGC
Segment 3	I6F688 DR4094	19 20	GCTGGCTCTATGGGTTTTGAAGT CTGGAGCGGGGAGGTGGCCGTATT
***SULT1A3* Gene-Specific Amplifications**			
Long PCR	TL F2 TL R3	21 22	AATGCCCGCAACAGTGCCTGCTGCATAGAG ACGCTGCCCGGCGGACTCGACGTCCTCCACCATCTT
Segment 1	I1AF1329 I4R171	23 24	GAGAATCCCACTTTCTTGCTGTT GGGAACAGTCTATGCCACCATAC
Segment 2	I4F1308 I6R240	25 26	GGTTCCTCCTTTGCCAGTTCAAC GGACTAAGTATCTGATCCGTGG
Segment 3	I6F405 DR3666	27 28	GGGCCCCAGGGGTTGAGGCTCTT ATATGTGGCCCCACCGGGCATTC

1.4 COS-1 Cell Expression

Seven different SULT expression constructs were used to transfect COS-1 cells. These constructs included cDNA sequences for all of the common *SULT1A1* and 1A2 allozymes observed during the present experiments, 1A1*1, 1A1*2, 1A1*3, 1A2*1, 1A2*2, and 1A2*3, as well as *SULT1A3*. As a control, transfection was also performed with expression vector that lacked an insert. All SULT cDNA sequences used to create the expression constructs had either been cloned in our laboratory (*SULT1A1*2*, *SULT1A2*2*, *SULT1A3*), were obtained from the Expressed Sequence Tag (EST) database and American Type Culture Collection (*SULT1A1*3*, *SULT1A2*1*) or were created by site directed mutagenesis (*SULT1A1*1*, *SULT1A2*3*). Each SULT cDNA was then amplified with the PCR and was subcloned into the eukaryotic expression vector pCR3.1 (Invitrogen, San Diego, CA). All inserts were sequenced after subcloning to assure that no variant sequence had been introduced during the PCR amplifications. COS-1 cells were then transfected with these expression constructs by use of the DEAE-dextran method. After 48 hr in culture, the transfected cells were harvested and cytosols were prepared as described by Wood, T.C. et al., Biochem. Biophys. Res. Commun., 198:1119-1127 (1994). Aliquots of these cytosol preparations were stored at -80°C prior to assay.

1.5 Data Analysis

Apparent K_m values were calculated by using the method of Wilkinson with a computer program written by Cleland. Wilkinson, G.N., Biochem. J., 80:324-332 (1961); and Cleland, W.W., Nature, 198:463-365 (1963). IC_{50} values and 50% thermal inactivation (T_{50}) values were calculated with the GraphPAD InPlot program (GraphPAD InPlot Software, San Diego, CA). Statistical comparisons of data were performed by ANOVA with the StatView program, version 4.5 (Abacus Concepts, Inc., Berkeley, CA). Linkage analysis was performed using the EH program developed by Terwilliger

and Ott, Handbook of Human Genetic Linkage, The Johns Hopkins University Press, Baltimore, pp. 188-193 (1994).

2.0

The experiments were performed in an attempt to identify common variant alleles for *SULT1A1* and *SULT1A2*, to determine the biochemical and physical properties of allozymes encoded by common alleles for *SULT1A2* and *SULT1A1* and to determine whether those alleles might by systematically associated with variation in TS PST phenotype in an important drug-metabolizing organ, the human liver. To achieve these goals, a stepwise strategy was utilized that took advantage of the availability of a "bank" of human hepatic biopsy samples which could be phenotyped for level of TS PST activity and thermal stability. DNA sequence information was available for each of the three known human PST genes (*SULT1A1*, *SULT1A2* and *SULT1A3*). *SULT1A1* and *SULT1A2* are located in close proximity within a 50 kb region on human chromosome 16. Raftogianis, R. et al., Pharmacogenetics, 6:473-487 (1996).

All exons for both *SULT1A1* and *SULT1A2* were sequenced using DNA from 150 platelet samples and 61 hepatic tissue samples to detect nucleotide polymorphisms and to determine whether there were significant correlations between genotypes for *SULT1A2* and/or *SULT1A1* and TS PST phenotype.

2.1 SULT1A2 and SULT1A1 Genetic Polymorphisms

All exons encoding protein for both *SULT1A2* and *SULT1A1* were PCR amplified in three segments (Fig. 2), and were then sequenced on both strands. Approximately 2 kb of DNA was sequenced for each gene. Therefore, a total of approximately 300 kB and 250 kB of sequence was analyzed for the 150 platelet samples and 61 hepatic biopsy samples, respectively. Thirteen different *SULT1A2* alleles were observed among the 122 alleles sequenced in the 61 biopsy samples. These alleles resulted from various combinations of ten different single nucleotide polymorphisms

(SNPs) (Table 2A). In Table 2A, numbers at the top indicate the nucleotide position within the ORF, in which 1= the "A" in the "ATG" start codon; or introns, in which an "I" followed by a numeral indicates the location of the nucleotide within the intron (i.e., I2-34 is the 34th nucleotide from the 5′-end of intron 2). Nucleotides shown as white type against a black background alter the encoded amino acid. Nucleotides 895 and 902 lie within the 3′-UTR of the *SULT1A2* mRNA. The values shown in the right-hand column indicate allele frequencies in the 61 hepatic biopsy samples.

Four of the *SULT1A2* SNPs altered the encoded amino acid, resulting in six different *SULT1A2* allozymes, three of which appeared to be "common" (frequency $\geq 1\%$, Table 2B). In Table 2B, numbers at the top indicate amino acid position from the N-terminus. The right-hand column indicates allozyme frequencies in the 61 hepatic biopsy samples studies. The other three alleles were observed only once, but their existence was confirmed by independent PCR and sequencing reactions. The allele nomenclature used here assigns different numerals after the * to alleles that encode different allozymes, with a subsequent alphabetic designation for alleles that also differ with regard to "silent" SNPs. Since population data was obtained, numeric assignments were not made randomly, but rather could be assigned on the basis of relative allele frequency in the population sample studied, i.e., **1* was more frequent than **2*, **2* was more common than was **3*, etc.

TABLE 2A
SULT1A2 ALLELES

	Exon II					Exon VI	Exon VII		Exon VIII		Allozyme Frequency 61 Hepatic Biopsy Samples
	20	24	56	I2-34	I5-78	506	704	I7-9	895	902	
*1A	T	T	C	T	T	C	A	C	T	A	0.467
*1B	T	T	C	T	C	C	A	C	T	A	0.025
*1C	T	T	C	C	C	C	A	C	T	A	0.008
*1D	T	T	C	T	C	C	A	C	C	A	0.008
*2A	C	C	C	C	C	C	C	C	C	G	0.262
*2B	C	C	C	T	C	C	C	C	C	G	0.016
*2C	C	C	C	C	C	C	C	T	C	G	0.008
*3A	T	T	T	T	C	C	A	C	T	A	0.156
*3B	T	T	T	T	T	C	A	C	T	A	0.016
*3C	T	T	T	T	C	C	A	T	T	A	0.008
*4	C	C	C	C	C	T	C	C	C	G	0.008
*5	C	C	C	C	C	C	A	C	C	G	0.008
*6	T	T	C	T	T	C	C	C	C	G	0.008

TABLE 2B
***SULT1A2* ALLOZYMES**

Allozyme	Amino Acid 7	19	184	235	Allozyme Frequency 61 Hepatic Biopsy Samples
*1	Ile	Pro	Arg	Asn	0.508
*2	Thr	Pro	Arg	Thr	0.287
*3	Ile	Leu	Arg	Asn	0.180
*4	Thr	Pro	Cys	Thr	0.008
*5	Thr	Pro	Arg	Asn	0.008
*6	Ile	Pro	Arg	Thr	0.008

Thirteen different *SULT1A1* alleles were detected in the platelet samples. These alleles encoded four different allozymes for *SULT1A1* (Table 3). In Table 3, numbers at the top indicate the nucleotide position within the ORF, in which 1=the "A" in the "ATG" start codon; or introns, in which an "I" followed by a numeral indicates the intron number, and the number after the dash indicates the location of the nucleotide within the intron (i.e., I5-34 is the 34th nucleotide from the 5'-end of the 5th intron). Nucleotides 902 and 973 lie within the 3'-UTR of the *SULT1A1* mRNA. The values in the right-hand columns indicate allele frequencies in the 61 hepatic biopsy samples studied or in DNA from 150 randomly selected Caucasian blood donors.

The 61 liver samples contained 10 of the 13 *SULT1A1* alleles identified in platelets, and encoded three of the four *SULT1A1* allozymes. Alleles *SULT1A1**1G*, **1H*, **1I*, **3A* and **4* were not present in these liver samples, but two novel *SULT1A1* alleles, **1J* and **1K*, were detected, bringing the total number of *SULT1A1* alleles identified to fifteen. These fifteen alleles involve various permutations of 24 individual SNPs located within the approximately 2 kb of *SULT1A1* DNA sequenced (Table 4).

TABLE 3
SULT1A1 SNP

SULT1A1 ALLELES																											
Allele	I1A-I38	Exon II		Exon III												Exon VII							Exon VIII		Allele Frequency 61 Hepatic Biopsy Samples	Allele Frequency 150 Random Blood Donors
		57	110	153	162	I5-34	I5-35	I6-11	I6-14	I6-17	I6-35	I6-45	I6-64	I6-488	I6-509	600	638	645	667	I7-16	I7-69	I7-120	902	973		
*1A	T	G	G	T	A	C	G	C	T	A	A	C	A	T	G	G	G	G	A	C	T	C	A	C	0.328	0.303
*1B	T	G	G	T	A	C	A	G	C	T	T	A	G	T	G	G	G	G	A	C	T	C	A	C	0.221	0.237
*1C	T	A	G	C	G	C	G	C	T	A	A	C	A	C	A	C	G	A	A	C	C	C	A	T	0.041	0 040
*1D	C	G	G	T	A	C	G	C	T	A	A	C	A	T	G	G	G	G	A	C	T	C	A	C	0.016	0.027
*1E	T	G	G	T	A	C	A	G	C	T	T	A	G	T	G	G	G	G	A	C	T	G	A	C	0.016	0 020
*1F	C	G	G	T	A	C	A	G	C	T	T	A	G	T	G	G	G	G	A	C	T	C	A	C	0.033	0.017
*1G	T	G	G	T	A	C	G	G	C	T	T	A	G	T	G	G	G	G	A	C	T	C	A	C	N.D.	0.010
*1H	T	A	G	C	G	C	G	C	T	A	A	C	A	T	G	G	G	G	A	C	C	C	A	C	N.D.	0.010
*1I	T	A	G	T	A	C	G	C	T	A	A	C	A	T	G	G	G	G	A	C	T	C	A	C	N.D.	0.007
*1J	T	A	G	C	G	C	G	C	T	A	A	C	A	C	A	C	G	G	A	C	C	C	G	T	0.008	N.D.
*1K	T	G	G	C	G	C	G	C	T	A	A	C	A	C	A	C	G	A	A	C	C	C	A	C	0.008	N.D.
*2	T	G	G	C	G	C	G	C	T	A	A	C	A	C	A	C	A	G	A	T	C	C	G	T	0.311	0.313
*3A	T	G	G	T	A	C	G	C	T	A	A	C	A	T	G	G	G	G	G	C	T	C	A	C	N.D.	0.007
*3B	T	G	G	T	A	T	A	G	C	T	T	A	G	T	G	G	G	G	G	C	T	C	A	C	0.016	0.003
*4	T	A	A	T	A	C	A	G	C	T	T	A	G	T	G	G	G	G	A	C	T	C	A	C	N.D.	0.003

TABLE 4
SULT1A1 ALLOZYMES

Allozyme	Amino Acid 37	213	223	Allozyme Frequency 61 Hepatic Biopsy Samples	Allozyme Frequency 150 Random Blood Donors
*1	Arg	Arg	Met	0.671	0.674
*2	Arg	His	Met	0.311	0.313
*3	Arg	Arg	Val	0.016	0.010
*4	Gln	Arg	Met	N.D.	0.003

The newly discovered alleles for *SULT1A2* appeared to be in linkage disequilibrium with alleles for *SULT1A1*. *SULT1A1*1* and **3* were linked to *SULT1A2*1* and **3* while *SULT1A1*2* was linked to *SULT1A2*2*. In this analysis, the hypothesis of no association between the two polymorphisms was rejected, but the hypothesis of association was supported with $x^2 = 53.83$ ($p < 0.0001$). Of the 122 sets of *1A1/1A2* alleles sequenced for each gene, only ten displayed discordance. The linkage disequilibrium complicated attempts to determine which of these two gene products might be responsible for phenol SULT phenotype. Therefore, to clarify possible genotype-phenotype correlations for these enzymes, biochemical and physical properties of the proteins encoded by all common alleles for *SULT1A1* and *SULT1A2* were determined.

2.2 COS-1 Cell Expression of SULT1A1 and SULT1A2 Allozymes

Expression constructs for each of the common (frequencies ≥ 1%) allozymes for *SULT1A1* and *SULT1A2* were used to transfect COS-1 cells. Selected biochemical and physical properties of the expressed enzymes were then determined. Those properties included apparent K_m values for the two cosubstrates for the enzyme reaction (4-nitrophenol and PAPS); thermal stability; and sensitivity to inhibition by DCNP. The substrate kinetic experiments were performed in two steps. Initially a wide range of concentrations of 4-nitrophenol that varied over at least three orders of

magnitude was tested, followed by detailed study of concentrations close to the apparent K_m value for that allozyme. Concentrations of 4-nitrophenol that were used to calculate apparent K_m values ranged from 0.02 to 5.0 μM for *SULT1A1*1*, 1A1*2 and 1A1*3; 0.08 to 10.0 μM for *SULT1A2*1* and 1A2*3; 1.0 to 1000 μM for *SULT1A2*2*; and 3.9 to 3000 μM for *SULT1A3*. Data from these experiments were then used to construct double inverse plots that were used to calculate apparent K_m values (Table 5). The results of the substrate kinetic studies suggested that TS PST phenotype in human liver might be due primarily to the expression of *SULT1A1*, since optimal conditions for the assay of TS PST activity in the human liver involved the use of 4 μM 4-nitrophenol as a substrate. See, Campbell, N.R.C. et al., Biochem. Pharmacol., 36:1435-1446 (1987). This concentration would be optimal for assay of the activities of allozymes encoded by alleles for *SULT1A1*, but was below the apparent K_m values for all of the *SULT1A2* allozymes. Of particular importance for the genotype-phenotype correlation analysis described subsequently is the fact that *SULT1A2*2* has a very high apparent K_m value for 4-nitrophenol (Table 5).

Apparent K_m values of the recombinant SULTs for PAPS were also determined. In those studies, as well as in the thermal stability and DCNP inhibition experiments, the concentrations of 4-nitrophenol used to perform the assays were 4 μM for SULT1A1*1, *2, and *3; 100 μM for SULT1A2*1; 50 μM for 1A2*3; and 3000 μM for SULT1A2*2 and SULT1A3. These concentrations were based on results of the 4-nitrophenol substrate kinetic experiments and represented the concentration at which maximal activity had been observed for that particular allozyme. Apparent K_m values of the recombinant SULT proteins for PAPS are also listed in Table 5. With one exception, those values varied from approximately 0.2 to 1.2 μM. The single exception was SULT1A2*1, with an apparent K_m value approximately an order of magnitude lower than those of the other enzymes studied (Table 5). Each value in Table 5 represents the mean ± SEM of nine separate determinations.

The thermal stabilities of the seven expressed proteins were also determined and varied widely. The rank order of the thermal stabilities was 1A2*2 > 1A2*1 >>

1A1*1 ≃ 1A1*3 ≃ 1A2*3 > 1A1*2 >> 1A3 (Table 5). These observations were consistent with experiments described herein that indicated that SULT1A1*2 was associated with a "thermolabile" phenotype in the platelet (Fig. 1) since that allele had the lowest T_{50} value of the recombinant "TS-PST-like" allozymes studied (Table 5). It is unlikely that allozyme SULT1A2*2 could explain a "thermolabile" phenotype since it was the most "thermostable" of the allozymes studied.

Finally, sensitivity of the recombinant proteins to inhibition by DCNP was determined. Sixteen different concentrations of DCNP, ranging from 0.01 to 1000 μM, were tested with each recombinant allozyme. IC_{50} values for DCNP also varied widely, with SULT1A2*3 being most, and SULT1A3 least sensitive to inhibition (Table 5). After all of these data had been obtained, the final step in this series of experiments was an attempt to correlate human liver TS PST phenotype with *SULT1A1* and/or *SULT1A2* genotype.

TABLE 5

RECOMBINANT HUMAN SULT BIOCHEMICAL AND PHYSICAL PROPERTIES

Allozyme	Apparent Km (μM)		Thermal Stability	DCNP Inhibition
	4-Nitrophenol	PAPS	T_{50} (°C)	IC_{50} (μM)
SULT1A1				
*1	0.88 ± 0.07	1.21 ± 0.02	39.3 ± 0.64	1.44 ± 0.11
*2	0.78 ± 0.08	0.98 ± 0.03	37.2 ± 0.43	1.38 ± 0.28
*3	0.31 ± 0.01	0.17 ± 0.02	38.9 ± 0.03	1.32 ± 0.27
SULT1A2				
*1	8.70 ± 1.10	0.05 ± 0.001	43.6 ± 0.15	6.94 ± 0.55
*2	373 ± 33	0.50 ± 0.001	46.3 ± 0.09	44.4 ± 1.50
*3	5.65 ± 1.14	0.28 ± 0.006	38.8 ± 0.19	0.97 ± 0.001
SULT1A3	4960 ± 810	0.28 ± 0.001	32.6 ± 0.19	86.9 ± 6.00

2.3 Human Liver Genotype-Phenotype Correlation

TS PST activity and thermal stability was measured in human platelet samples (n=905) and human liver biopsy samples (n=61). A scattergram of these data are shown in Figure 1 and 2. Subjects homozygous for the allele *SULT1A1*2* uniformly had low levels of both TS PST activity and thermal stability in their platelets (Fig. 1B). The genotype-phenotype correlation for *SULT1A1* in the liver samples is shown in Fig. 4A. Similar data for *SULT1A2* are plotted in Fig. 4B. Figure 4 demonstrates that the *SULT1A1*2* allele appeared to be associated with low TS PST thermal stability in the liver, just as it was in the human blood platelet (Fig. 1B). For example, the average H/C ratio for samples homozygous for *SULT1A1*1* was 0.57 ± 0.01 (n=28, mean ± SEM), while that for heterozygous *1A1*1/1A1*2* samples was 0.40 ± 0.01 (n=24) and that for samples homozygous for *SULT1A1*2* was 0.18 ± 0.01 (n=7, $p < 0.001$ by ANOVA). Table 6 summarizes this data.

TABLE 6
SULT1A1 ALLOZYMES AND TS PST ACTIVITY

Platelet Allozyme	AA 213	N	H/C Ratio	N	TS PST activity
*1/*1	Arg/Arg	11	0.62±0.03**	11	1.08±0.25
*1/*2	Arg/His	8	0.53±0.03**	9	0.90±0.20
*2/*2	His/His	13	0.09±0.02**	13	0.14±0.01
Liver					
*1/*1	Arg/Arg	28[a]	57.5±1.31*	28[a]	56.0±3.05
*1/*2	Arg/His	24	40.5±1.38*	24	56.8±4.19
*2/*2	His/His	7	17.7±1.44*	3[b]	28.5±2.27**

* $p<0.0001$ by ANOVA compared with other two groups; **$p<0.02$ by ANOVA compared with other two groups; [a] Two samples heterozygous for *SULT1A*3* were not

included in these analyses; [b] Four malignant hepatic samples homozygous for *SULT1A1*2* were not included in this analysis.

Although the *SULT1A1*2* allele was highly correlated with low TS PST thermal stability in the liver, unlike the situation in the platelet, low thermal stability was not significantly correlated with low levels of TS PST activity (Fig. 4A). Of possible importance is the fact that, when the data were stratified on the basis of diagnosis, of the seven samples homozygous for *SULT1A1*2*, the three from patients with benign hepatic disease had the lowest levels of TS PST activity, while the four samples from patients with malignant disease had the highest activity (28.5 $\pm$ 2.3 vs. 59.8 $\pm$ 4.0, mean $\pm$ SEM respectively, $p < 0.002$).

The results of the substrate kinetic experiments (Table 5), as well as the results of the thermal stability studies suggested that TS PST phenotype in the liver was most likely a measure of *SULT1A1* expression. As pointed out previously, that was true because both K_m values for 4-nitrophenol and T_{50} values for recombinant SULT1A2 allozymes were above those found to be optimal for the determination of TS PST phenotype in human liver cytosol preparations (Table 5). Testing that hypothesis directly is complicated by the fact that *SULT1A1* and *1A2* share 95% or greater identity for both protein amino acid and mRNA nucleotide sequences; so neither Western nor Northern blots can easily distinguish between them. However, biochemical studies of recombinant SULT allozymes suggested that the sulfation of 100 μM 4-nitrophenol might represent a relatively specific measure of SULT1A2 activity (Table 5). As a result of the profound substrate inhibition which these enzymes display, SULT1A1 allozymes show little or no activity at that concentration, and SULT1A3 would not contribute significantly to activity measure at that concentration because of its very high K_m value for 4-nitrophenol (Table 6). Therefore, 100 μM 4-nitrophenol was used as a substrate with cytosol from six pooled liver samples in an attempt to measure SULT1A2 activity. However, after

three attempts no activity was detected, suggesting that SULT1A2 is not highly expressed in the liver. Ozawa, S. et al., Chem. Biol. Interact., 109:237-248 (1998).

In summary, common genetic polymorphisms were observed for both *SULT1A1* and *SULT1A2* in humans. However, the proteins encoded by these alleles differed in their biochemical and physical properties. Recombinant SULT1A2*2 had a K_m value dramatically higher than did SULT1A2*1 or 1A2*3. The allele *SULT1A1*2* was associated with decreased TS PST thermal stability in the liver and in the blood platelet. Unlike the situation in the platelet, *SULT1A1* or *SULT1A2* alleles identified in the hepatic tissues did not appear to be systematically associated with level of TS PST activity.

2.4 SULT1A3 Polymorphisms

All exons and introns for *SULT1A3* were sequenced using DNA from 150 random blood donor samples to detect nucleotide polymorphisms. Table 7 describes sequence variants.

TABLE 7
Nucleotide Transition/Transversion and Position Within SULT1A3 Gene

Classification	Exon 3 105	I3-83/84 Insertion	I4-69	I6-69	I7-113
Wild Type	A	---	G	G	G
Variant	G	CAGT	A	A	T

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of

the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.